



Scientext comme corpus pour l'enseignement

Cristelle Cavalla, Mathieu Loiseau

► To cite this version:

Cristelle Cavalla, Mathieu Loiseau. Scientext comme corpus pour l'enseignement. Agnès Tutin; Francis Grossmann. L'écrit scientifique: du lexique au discours. Autour de Scientext, Presses Universitaires de Rennes, pp.163-182, 2013. hal-01154754

HAL Id: hal-01154754

<https://hal.science/hal-01154754>

Submitted on 22 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SCIENTEXT COMME CORPUS POUR L'ENSEIGNEMENT

Cristelle Cavalla, Mathieu Loiseau

1 Introduction

Dans cet article, nous abordons la question de l'utilisation des corpus sur un terrain connu : l'enseignement et la recherche universitaire. Les travaux de recherche sur l'exploitation des corpus en France suggèrent qu'il s'agit d'une piste à même d'offrir des contributions positives à la didactique des langues. Cependant, comme Chambers (Chambers A., 2007), nous constatons que ces outils restent modérément utilisés en dehors du contexte de l'enseignement supérieur.

Notre contribution vise à proposer, dans le contexte de la didactique du Français Langue Etrangère, un prototype dans le cadre du projet FULS¹ (Formes et Usages des Lexiques Spécialisés pour l'enseignement en Français Langue Étrangère) qui utilise le corpus Scientext. À ce titre – celui du « détournement » d'un instrument prévu pour une utilisation dans la recherche en linguistique – le corpus Scientext sera manipulé dans le cadre de l'enseignement des langues et plus particulièrement du FLE/FOU². Le FOU (Mangiante J.-M. et Parpette C., 2011) est une spécialité désormais reconnue dans le domaine du FLE pour aider les apprenants-étudiants en FLE à mieux comprendre le monde universitaire français et donc à mieux s'intégrer dans cette institution avec ses règles et ses codes spécifiques. La séquence présentée ici est centrée sur l'un des aspects linguistiques du FOU : la phraséologie scientifique. A la suite d'une étude sur des écrits d'étudiants locuteurs non natifs (LNN) dont nous avons observé les lacunes, nous avons proposé des activités sur ce sujet. Nous introduirons ensuite la séquence d'apprentissage proposée dans FULS, pour conclure sur les perspectives offertes par ce travail.

2 Les corpus dans l'enseignement des langues

Si l'utilisation des corpus pour la description linguistique est désormais acquise, il en va tout autrement pour l'enseignement des langues étrangères. Malgré les avancées linguistiques mises en évidence grâce à l'analyse des corpus oraux et écrits, l'enseignement des langues ne s'est pas encore véritablement approprié les outils et les corpus, notamment l'enseignement du FLE. Notons toutefois que les corpus oraux sont davantage entrés dans les pratiques pédagogiques que les corpus écrits.

Les corpus, en particulier les corpus d'apprenants, commencent à bénéficier d'une plus grande popularité dans les recherches en didactique des langues³, car d'après les linguistes didacticiens, ils fournissent des bases « empiriques beaucoup plus solides que celles qui émanent des seules données introspectives » (Granger S., 2001). En outre, les grands corpus

¹Projet financé par le Ministère de la Recherche dans le cadre des « Programmes Pluri-Formation ». Projet dirigé au Lidilem par C.Cavalla de 2007 à 2010. Je remercie les personnes qui ont contribué au développement du projet et notamment Arnaud Bey (ingénieur d'études) sans qui le site et une partie des exercices de FULS n'existerait pas.

²Français Langue Étrangère / Français sur Objectifs Universitaires.

³Bien qu'ils soient encore peu utilisés dans les classes de langue en France comme le constate très justement Boulton (Boulton, 2008).

permettent de repérer des régularités et/ou des exceptions dans la langue (Blanche-Benveniste C., 1996) ce qui est primordial dans l'enseignement d'une langue étrangère (notamment dans le choix des lexies à enseigner). Enfin, rappelons qu'un corpus est d'autant plus pertinent qu'il est annoté utilement (même si l'annotation ne pourra jamais entièrement répondre à nos attentes en termes d'analyses linguistique (Habert B., 2004).

Dans les pays anglo-saxons, l'utilisation des corpus en didactique n'est pas nouvelle. Le développement de l'utilisation des corpus en didactique des langues s'est effectué dans les années 80 avec notamment le DDL (Data Driven Learning ou apprentissage guidé par les données) de Johns (Johns T., 1986) et l'entrée des concordanciers dans les classes de langue. Des manifestations scientifiques sont entièrement consacrées à l'exploitation didactique des corpus (conférence TALC depuis 1994 (<http://talc8.isla.pt/previoustalcs.html>)), et à leur constitution à visée lexicographiques, linguistiques, informatiques et autres (Péry-Woodley M.-P., 1995) ; reconnaissons cependant que ces manifestations suscitent encore peu d'intérêt dans la communauté des didacticiens français.

L'entrée des corpus en classe de langue pose la question de l'autonomie de l'apprenant et soulève des interrogations auxquelles les chercheurs commencent à répondre (Chambers A., 2005, 2010/2, Chambers A. et O'Sullivan I., 2004, Rézeau J., 2007, Schaeffer-Lacroix E., 2008) : les apprenants sont-ils à même d'analyser des corpus ou de s'accommoder d'annotations destinées à des linguistes ? Les usages des corpus proposés par Fligelstone (Fligelstone S., 1993) apportent un peu de lumière sur cette question.

...2.1 Trois points de vues sur les corpus

Selon Fligelstone (Boulton A., 2008, Fligelstone S., 1993), trois points de vues didactiques sont envisageables pour l'utilisation des corpus⁴ : « teaching about » (le corpus comme objet d'enseignement), « exploiting to teach » (le corpus comme support d'enseignement) et « teaching to exploit » (le corpus pour enseigner et apprendre la langue). Le premier usage, « teaching about », correspond à l'enseignement de la linguistique de corpus ou de sa méthodologie. Le deuxième usage, « teaching to exploit » correspond à l'enseignement de l'usage d'un corpus : quelles requêtes faire, dans quelle situation en fonction du système d'interrogation disponible ? Enfin « exploiting to teach » revient à exploiter les corpus et l'interprétation des résultats pour enseigner une langue (une L2 dans les cas qui nous intéressent ici).

Il est intéressant de noter que le premier type d'activités vise à développer chez les apprenants les connaissances générales nécessaires à mettre en place des schèmes d'usage des corpus (deuxième point de vue de Fligelstone). Ces derniers seront actifs dans la mise en œuvre d'un schème d'action instrumentée (utilisation d'un corpus en autonomie). Conscients de ces deux utilisations d'un corpus, nous voudrions les associer afin que les apprenants acquièrent simultanément des connaissances en langue étrangère et des compétences pour l'utilisation des corpus dans leur apprentissage.

...2.2 L'utilisation des corpus en classe de langue

Chambers et al. (Chambers A., Farr F. et O'Riordan S., 2011) remarquent que les enseignants manquent de temps pour s'investir réellement dans l'utilisation optimale de ces outils. Cela pourrait expliquer le constat de Boulton selon lequel « l'exploitation directe des corpus par les enseignants et surtout par les apprenants n'est pas aussi avancée que l'auraient laissé supposer les nombreux articles parus dans les années 90. » (Boulton A., 2008 : 37). Nous

⁴Fligelstone aborde dans son article le thème des corpus et de l'enseignement (CAT : *Corpus and Teaching*), mais pas uniquement dans le cadre de l'apprentissage des langues (Fligelstone, 1993 :102).

notons en effet qu'en France, les enseignants de FLE ont tendance à bouder l'utilisation des corpus. Ce constat peut être corrélé au fait que peu de corpus sont prévus pour une exploitation didactique directe des données. En d'autres termes, les corpus tels que Scientext, Frantext, ou quelques corpus journalistiques ou littéraires en ligne en accès libre (Lextutor par exemple), permettent l'extraction de concordanciers fort pertinents pour l'enseignement, mais ne proposent aucune activité ou scénario pédagogique à l'enseignant débutant. Ajoutons que les requêtes sont parfois complexes (Frantext notamment) ce qui peut rebuter certains utilisateurs (et rappelons le manque de temps des enseignants évoqué par Chambers).

La formation des enseignants de FLE entre également en jeu dans la méconnaissance des corpus par ces derniers. Les filières FLE universitaires françaises ne proposent que peu de cours présentant ces outils et donc les techniques pour exploiter les corpus en classe de FLE (ce qui est différent pour l'anglais langue étrangère (Maniez F., Dury P., Arlin N. et Rougemont C., 2008, Osborne J., 2008)). On trouve toutefois des enseignements de ce type dans quelques formations universitaires, à Avignon (Master FLE/FLES Éducation interculturelle) et à Bordeaux 3 (Master Professionnel Didactique FLES).

Les exemples d'utilisation des corpus montrent, pour la plupart, que les enseignants ne créent pas de nouveaux outils, mais tentent d'intégrer à leur classe ceux existants (Rabardel P., 1995). Détourner l'usage d'un outil reste moins chronophage et tout aussi rentable – apparemment – que le développement d'un nouvel outil ou de l'extension d'un outil existant. Par exemple, Aston (Aston G., 2002) considère qu'il est plus profitable (gain de temps notamment) pour un enseignant ou un chercheur d'utiliser un corpus existant avec les outils existants et d'en faire ce que bon lui semble plutôt que de fabriquer son propre corpus. D'après Chambers (Chambers A., Farr F. et O'Riordan S., 2011), les enseignants ont tendance à utiliser les technologies (outils d'interrogation des corpus en ligne) selon leurs habitudes pédagogiques sans modifier l'outil d'après leurs besoins, en s'adaptant à l'outil sans en maîtriser toutes les potentialités :

« In general 'teachers are using technology in ways that fit their current practice, rather than transforming their practice through the use of technology' » (Chambers et al., 2011: 87).

Ainsi, guider les enseignants (comme les apprenants) dans l'utilisation de ces outils, paraît-il pertinent mais il ne faut pas qu'ils aient l'impression de perdre du temps. De notre côté, l'utilisation que nous préconisons suit la progression didactique suivante : 1/ placer l'utilisateur face à des exercices élaborés à partir d'un corpus (Alfalex⁵ par exemple, même si le corpus utilisé est clairement présenté aux apprenants), 2/ entraîner l'utilisateur vers la découverte des ressources à disposition dans le corpus en question (consultation indirecte), 3/ conduire l'utilisateur vers la prise en main de l'outil d'interrogation du corpus. La première étape proposée est stratégiquement discutable : pourquoi placer l'utilisateur face à des exercices alors que nous voudrions qu'il utilise le corpus ? Chambers ainsi que Boulton ou Granger, préconisent de placer l'apprenant (pourquoi pas l'enseignant débutant également) devant de petits corpus dans un premier temps afin de ne pas le noyer d'informations. Nous pensons alors qu'il serait profitable de proposer des exercices qui utilisent les phrases, les concordances du corpus et donc qui plongent l'apprenant et l'enseignant dans le « bain linguistique » du corpus. Au plan discursif, la découverte des spécificités des écrits universitaires n'est pas évidente ; au plan pratique, le travail sur des phrases courtes ou de courts paragraphes évoquant des sujets inconnus de l'utilisateur⁶ n'est pas aisé. Une aide pour la découverte tant du lexique que des extraits proposés serait une étape non négligeable pour ensuite conduire l'apprenant et

⁵Alfalex (Grelep, Leuven, Belgique) : <http://www.kuleuven.be/alfalex/>

⁶Même si l'utilisateur de Scientext est linguiste et choisit le corpus en linguistique, il ne peut être spécialiste de toutes les branches de la discipline et donc ne pas connaître certaines notions.

l'enseignant vers l'utilisation du corpus en sachant ce qu'il peut y trouver et donc ce qu'il peut chercher. En effet, comment faire des requêtes quand on ne connaît pas le contenu du corpus ? L'abord par des exercices dont les extraits sont issus du corpus donne à voir – au futur utilisateur du corpus – le type d'extraction possible.

Le site Scientext propose des requêtes pour l'étude linguistique des écrits scientifiques et n'est pas a priori conçu pour une utilisation pédagogique : il faut donc envisager un détournement de l'outil pour l'exploiter en classe de langue. Comme il faut proposer un outil rapide à prendre en main, nous proposons cet outil en ligne : FULS. Ainsi, notre réflexion s'est-elle portée vers l'aide à l'enseignant pour entrer dans l'utilisation du corpus. L'idée – peut-être ambitieuse – est d'entraîner l'enseignant à découvrir le corpus en même temps que l'apprenant en suivant pas à pas la séquence didactique proposée à partir du corpus. Nous adoptons ainsi les deux dernières approches de Fligelstone de la façon suivante :

- « exploiting to teach » (le corpus comme support d'enseignement) : l'enseignant et l'apprenant suivent une séquence didactique développée à partir du corpus Scientext : nous exploitons les ressources offertes par le corpus pour enseigner un phénomène linguistique en langue étrangère. Ceci est proposé par plusieurs auteurs qui font travailler les apprenants sur des concordanciers soit prêts à l'emploi soit à construire en ligne ;
- puis « teaching to exploit » (le corpus pour enseigner et apprendre la langue) où l'apprenant s'approprie l'outil d'exploitation du corpus afin de le manipuler à sa guise.

La consultation directe permettra à l'apprenant de faire des recherches de mots ou d'expressions dans le corpus et de créer des concordances. Lors de la consultation indirecte (approche « indirecte » du corpus (Chambers A., 2010/2)), l'apprenant n'a pas accès au corpus lui-même, mais il consulte des concordances créées par l'enseignant (Chambers, 2010 : 15). Notons que nous n'envisageons pas seulement une recherche d'occurrences pour des exemples dans le corpus de la part de l'apprenant, mais une réflexion sur la langue tout en le contraignant à exploiter le corpus selon ses besoins linguistiques préalablement cernés à l'aide de l'enseignant.

...2.3 Pour l'acquisition de connaissances en L2

Pour la création d'activités linguistiques, nombre d'enseignants exploitent les concordanciers. Comme le signalent plusieurs auteurs (Flowerdew J., 1993, Johns T., 1986, McEnery T. et Wilson A., 1997, etc.), les concordances peuvent être utilisées en classe sous forme d'exempliers papier. Il s'agit dans ce cas-là principalement du troisième point de vue sur les corpus proposé par Fligelstone, c'est-à-dire l'utilisation directe du corpus par les apprenants de L2. Les enseignants utilisent les outils⁷ du système (proposé par l'outil d'interrogation du corpus) et instrumentalisent (Rabardel P., 1995) l'outil des linguistes pour générer des supports pédagogiques (qui peuvent dépasser l'exemplier, sous forme d'exercices lacunaires par exemple). Ils évacuent ainsi les difficultés techniques liées à l'usage du corpus dans la classe (besoins matériels, mise en place des activités, cf. « Learning about » / « Teaching to exploit » de Fligelstone), difficultés qui ne sont probablement pas étrangères à la faible utilisation dans la pratique d'enseignement des langues. Notons à ce propos que la plupart des expériences discutées dans Boulton (Boulton A., 2008: 41) permet aux apprenants de manipuler les corpus majoritairement dans des contextes universitaires. Les schèmes d'usages impliqués sont

⁷En général, uniquement le corpus et le système d'interrogation – concordancier – dans les exemples cités, mais rien n'interdirait de procéder de la même manière avec un corpus annoté.

principalement tournés vers l'outil « système d'interrogation » et l'interface avec les autres outils d'utilisation. Aston (Aston G., 2002: 10) recense cinq types d'activités (non exclusives) d'apprentissage des langues avec corpus :

1. activités centrées sur la forme (pour établir et pratiquer les usages de certains phénomènes linguistiques) ;
2. activités centrées sur le sens (influence du contexte sur le sens) ;
3. activités centrées sur un savoir-faire (compétences et stratégies de lecture, du type *bottom-up* (Brodine R., 2001)) ;
4. activités où les corpus servent de référence (tâches impliquant d'autres textes, le corpus sert d'aide à la lecture, l'écriture ou la traduction) ;
5. activités d'exploration (exploration spontanées des corpus).

Seuls les trois premiers types d'activités peuvent être réalisés à partir de supports papier. Les deux derniers nécessitent que les apprenants aient acquis une certaine autonomie dans l'interrogation et l'interprétation des résultats d'un corpus (schèmes d'usages). Le dernier point nous intéresse particulièrement car nous souhaiterions conduire les apprenants vers une utilisation autonome du corpus Scientext. L'exploration servirait à la fois pour l'aide à la rédaction, la découverte de nouvelles formes phraséologiques et autres types d'analyse linguistique.

...2.4 Compétences d'utilisation des corpus : apprenants et enseignants

Le recours à un corpus comme Scientext pour une séquence didactique est une démarche courante pour l'enseignement d'un lexique spécialisé en FLE. Johns proposait aux enseignants d'utiliser un concordancier sur les textes de leurs apprenants pour que ces derniers puissent comparer leurs usages de termes de spécialités avec ceux d'un petit corpus constitué d'articles du même type (Johns T., 1986: 160-161). Aston propose de laisser les apprenants constituer leur corpus et évoque notamment la créativité et/ou l'éveil de l'esprit critique par rapport au corpus (Aston G., 2002). Quand un enseignant crée son propre corpus, la taille (réduite) et l'homogénéité souvent inhérente à ce recueil de données (restreint), n'impliquent pas directement l'apprenant mais permet un accès plus rapide aux données. Aston invoque ainsi l'avantage offert par des corpus de petite taille qui permettent une plus grande facilité d'analyse, de prise en main, d'interprétation et de constitution, même s'ils perdent en représentativité (Aston G., 1997). Dans la même perspective, nous n'envisageons pas de mettre l'apprenant directement face au corpus Scientext qui est conséquent mais nous proposons un cheminement progressif vers la découverte du corpus dans son entier (voir Falaise ce volume).

Au sujet des grands corpus, Aston soulève une difficulté cruciale de leur intégration dans les classes. Si l'usage de grands corpus généralistes déjà constitués peut poser des problèmes du fait de la quantité d'information qu'ils contiennent, la constitution de corpus pour la pratique pédagogique est directement tributaire du système d'interrogation disponible. En termes davantage informatiques, nous pourrions voir si le système utilisé pour interroger le corpus peut fonctionner avec n'importe quel corpus. En effet, en règle générale, plus un système permet des interrogations précises, plus l'utilisateur est contraint en termes de données sur lesquelles travailler. Dans ce cas, en termes didactiques, nous questionnons davantage l'outil d'interrogation : est-il adapté aux besoins des apprenants ? La terminologie est-elle suffisamment claire pour que tout apprenant non spécialiste de linguistique puisse facilement l'utiliser ? Comment restreindre les données à extraire afin de ne pas noyer l'apprenant dans un flot d'exemples ?

Ainsi, malgré l'enthousiasme autour des corpus dans l'enseignement des langues, et la diversité des usages avérés, on se heurte à des difficultés dans la mise en place des activités

dans le choix des outils et les possibilités d'interrogation. Le développement d'une séquence autour de Scientext prend en compte ces éléments afin que les activités soient d'abord faites puis développées par les utilisateurs ; le corpus recèle des données pertinentes (au plan lexical, sémantique et syntaxique notamment) pour le public visé et les différents modes d'interrogation seront utilisés plus tard, une fois les apprenants à l'aise avec les notions linguistiques à rechercher. Nous pensons, par exemple, aux lexies fréquemment présentes dans les écrits universitaires et leurs associations lexicales. Les apprenants (et les enseignants) doivent les connaître pour les chercher et faire les requêtes en conséquence.

3 Utilisation du corpus Scientext pour FULS

Le projet FULS⁸ (Formes et Usages des Lexiques Spécialisés) propose d'impliquer les enseignants dans la découverte du corpus et de phénomènes linguistiques à enseigner aux apprenants, ces deux aspects s'effectuant alors simultanément. L'enseignant a la double tâche de découvrir le corpus en même temps que l'apprenant et de guider ce dernier dans l'apprentissage d'un phénomène linguistique en suivant les activités proposées.

Le français scientifique et universitaire a désormais un corpus accessible en ligne – Scientext – qui permet à tout enseignant et apprenant du français, de découvrir le lexique universitaire et scientifique. Scientext contient suffisamment de mots (près de 5 millions) pour que l'enseignement de FLE s'en empare à la fois pour l'enseignant et pour l'apprenant. Notre souhait à long terme est de rendre accessible à tous, en ligne, un outil d'aide à la rédaction d'écrits universitaires longs pour les étudiants étrangers venus suivre un cursus universitaire en Master et Doctorat en France (Cavalla C., 2010). Pour ce faire, nous utilisons le corpus et des outils de Traitement Automatique des Langues (TAL) et de Technologies de l'information et de la communication pour l'enseignement (TICE) pour une exploitation – nous l'espérons – optimale du corpus. Les questions qui nous guident sont de quatre ordres : d'abord, en termes linguistiques et didactiques, que peut-on envisager de faire chercher à un novice non natif dans un grand corpus en langue cible ? Est-il raisonnable de mettre des gens face à un outil complexe dans une langue qu'ils ne maîtrisent pas encore ? Puis, en termes d'utilisation, doit-on faire en sorte que l'utilisateur manipule le corpus ou qu'il recourt simplement à des séquences préconstruites ? Enfin, en termes techniques, le TAL peut-il apporter des outils que nous ne pouvons trouver ailleurs pour l'enseignement des langues ?

L'idée développée suit plusieurs recommandations précédemment évoquées : d'une part, pour l'enseignant, proposer de découvrir le corpus au fil d'une séquence d'apprentissage afin de lui faire gagner du temps et qu'il puisse ainsi davantage se concentrer à l'enseignement plutôt qu'à la découverte, parfois longue, de l'utilisation appropriée de nouveaux outils pédagogiques ; d'autre part, en ce qui concerne l'apprenant, éviter de trop grands corpus comme préconisé par Chambers ou proposer une utilisation progressive d'un grand corpus. Chambers précise qu'il faudrait soit le découper et n'afficher qu'une partie (et proposer une sélection), soit fabriquer son propre petit corpus (avec ou sans les apprenants) ce qui évitera aux apprenants de se perdre dans des exemples trop fournis ou trop complexes (le tri peut être fait à plusieurs niveaux). Nous proposons alors de guider l'enseignant et les apprenants (simultanément) vers le corpus sans le présenter préalablement.

...3.1 Utilité des corpus écrits en FLE pour la découverte de la phraséologie

Pour faire face aux difficultés des apprenants de Master et doctorat à rédiger leurs écrits

⁸Dirigé par C.Cavalla, Lidilem, 2007-2010. Financé par le Ministère de la Recherche dans le cadre des projets Plan Pluri-Formation. Description en ligne : <http://scientext.msh-alpes.fr/fuls/>, http://webtek-66.iut2.upmf-grenoble.fr/0_0_0_Accueil_0.html. Une bibliographie sur le site permet de plus amples détails.

universitaires scientifiques, Scientext constitue un outil qui peut les aider à découvrir certaines des attentes, en matière d'écrits, des universités françaises.

L'utilisation d'un corpus pour l'aide à la rédaction est associée à un objectif linguistique précis : l'enseignement / apprentissage de la phraséologie transdisciplinaire et plus précisément des collocations fréquentes rencontrées dans ce genre d'écrit. Stevens, comme Johns (2002) considère que l'apprentissage du lexique est très efficace par l'analyse des concordances (Chambers et al., 2011 : 88). En effet, la recherche de cooccurrences lexicales est facilitée par l'alignement des éléments dans un concordancier et permet une analyse lexicale, syntaxique et sémantique des éléments en question. L'extraction de phrases-exemples pour l'élaboration des séquences est simplifiée pour l'enseignant qui peut avoir accès à des listes de collocations que le linguiste met à sa disposition. Les exemples sont alors multipliés et peuvent être générés automatiquement à partir du corpus. Ceci permet de proposer le même format d'exercice mais avec des exemples de travail différents à chaque nouveau passage.

Nous rencontrons ce lien corpus/générateur d'activités sur un site ouvert à tous : Alfalex (Verlinde S., Binon J. et Selva T., 2006). En lien avec un corpus journalistique, un corpus d'écrit de la Communauté Européenne analysés avec des outils de TAL et un dictionnaire (la BLF : Base Lexicale du Français) (Verlinde S., Selva T. et Binon J., 2009), ce site génère des exercices essentiellement lacunaires dont les exemples sont sans cesse renouvelés automatiquement grâce au lien avec les corpus.

Nous souhaitons développer ce genre d'interaction entre FULS (site d'activités) Scientext et l'utilisateur. L'utilisateur a accès, via FULS, à la diversité textuelle et disciplinaire proposée dans Scientext, ce qui permet à l'étudiant, novice dans le monde universitaire français, de découvrir les différentes formes d'écrits exigés par cette institution. En outre, Scientext propose un nombre conséquent de textes récents en accès libre, ce qui permet le choix et donc le tri des formes lexicales recherchées. Enfin, Scientext est un corpus annoté morphologiquement et syntaxiquement à l'aide de l'analyseur syntaxique Syntex (Bourigault D., Fabre C., Frérot C., Jacques M.-P. et Ozdowska S., 2005). Des grammaires sémantiques, par exemple les 'verbes d'opinion' exploitant les informations lexicales et syntaxiques, sont proposées et peuvent être d'ores et déjà exploitées par les utilisateurs. Ceci sera pris en compte ultérieurement pour l'enseignement du FLE dans FULS.

...3.2 La séquence FULS

Le projet FULS, qui conduit les apprenants et les enseignants vers l'utilisation d'un corpus, peut être envisagé comme une application de Scientext pour la didactique du FLE/FOU (Cavalla C., 2008). Dans un premier temps, seules les collocations (Binon J. et Verlinde S., 2003, Blumenthal P. et Hausmann F. J. e., 2006, Tutin A. et Grossmann F., 2002) transdisciplinaires sont abordées. Scientext renferme ces expressions spécifiques aux écrits universitaires de plusieurs disciplines, comme *formuler une hypothèse* ou *résultats encourageants* (Tutin A., 2007, 2007). Ces éléments sont alors extraits pour faire des exercices développés grâce à des techniques du TAL et un outil TICE d'aide à la rédaction présentés ci-dessous. Les exercices du projet FULS apparaissent dans un site Internet dont l'accès est encore limité car il est toujours en cours d'élaboration⁹. Parmi les différents domaines du FOU, la séquence FULS apparaît dans l'aide à l'écrit universitaire au plan linguistique uniquement (nous n'aborderons pas la méthodologie dans cette séquence).

Les collocations proposées dans la séquence appartiennent à des champs sémantiques spécifiques étudiés dans le cadre du projet Scientext. Par exemple, nous avons extrait et étudié les collocations désignant la démarcation de l'auteur par rapport à ses pairs ou à une théorie

⁹Le site devrait bientôt être en accès libre à l'adresse suivante : scientext.msh-alpes.fr/fuls/prototypeFULS.html

(par exemple, *contrairement à X,...nous nous distinguons de X*), les expressions concernant l'évaluation de son propre travail ou du travail d'autrui (ex: *méthode prometteuse, hypothèse valide*). Les collocations choisies appartenant à un champ sémantique spécifique, nous avons choisi de les présenter en les associant à des fonctions rhétoriques et sémantiques afin de guider l'apprenant dans son écriture (on évalue davantage en conclusion que dans la partie théorique ou dans la partie descriptive). L'articulation de ces spécificités a permis d'élaborer une séquence didactique pour l'enseignement de ces éléments.

...3.2.1 Progression de la séquence

La séquence FULS suit une progression classique de la découverte du phénomène vers la réutilisation personnelle (pour appropriation) des éléments en question. Trois grandes étapes complémentaires s'articulent et se complètent au fil de la séquence : 1/ Découverte des collocations transdisciplinaires scientifiques ; 2/ Approfondissement des savoirs par entrées sémantiques et donc utilisation de collocations spécifiques ; 3/ Réemploi en contexte rédactionnel de toutes les collocations abordées. Le découpage par champs sémantiques permet de situer l'apprenant à un moment de son écriture et donc l'entraîne à l'utilisation de collocations spécifiques. Toutes les phrases des exercices sont extraites du corpus Scientext. L'apprenant travaille hors corpus (approche indirecte du corpus (Chambers A., 2010/2)), sur des extraits qui changent à chaque utilisation pour la plupart des séquences, même si certaines collocations et exemples d'exercices sont répétés pour faciliter la mémorisation. La figure 1 résume l'articulation didactique de ces différents aspects.

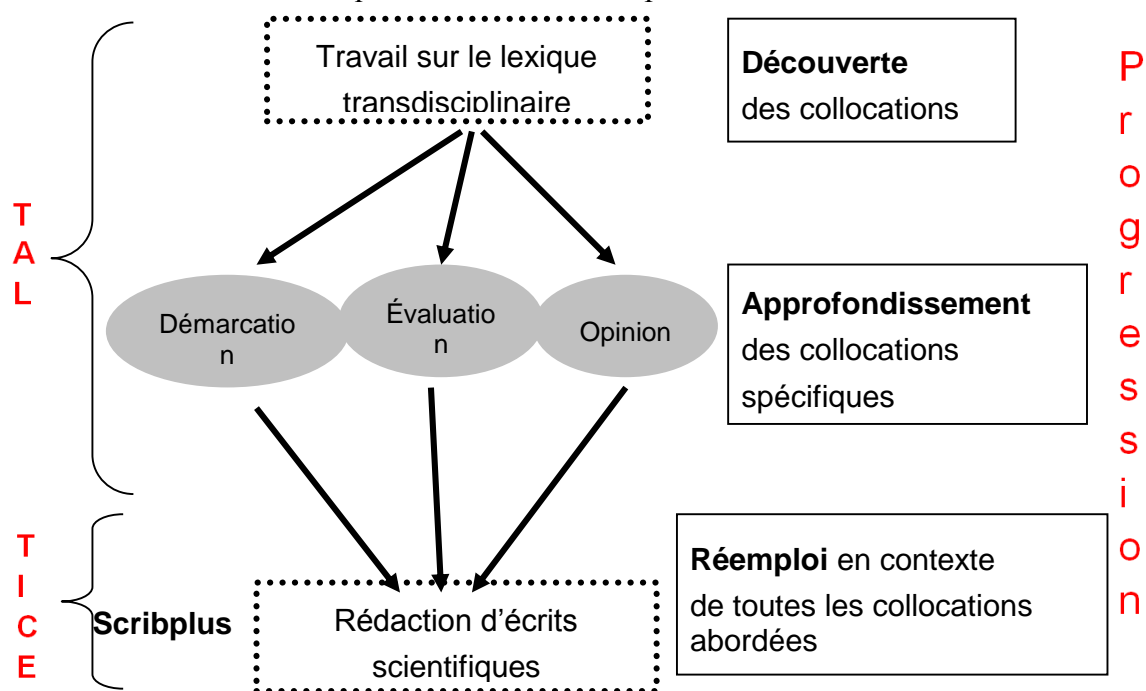


Figure 1 : Articulation des éléments spécifiques pour l'élaboration de la séquence FULS

La figure 1, lue de haut en bas, permet de suivre la progression didactique de la séquence. Le site permet à l'apprenant de choisir un champ sémantique et donc de ne pas suivre l'ordre proposé. La figure 2, sur laquelle on peut voir la page d'accueil du site, propose, à gauche de l'écran, la séquence « Découverte » qui rassemble toutes les collocations de tous les champs sémantiques, puis les champs sémantiques que l'apprenant choisit. Ceci est expliqué à l'utilisateur sur cette page dans la rubrique « Comment faire les exercices ? ».



séquence :

- Il est d'abord proposé un exercice lacunaire assez classique qui tente d'aider à la mémorisation des collocations. Il est proposé de réitérer trois fois les mêmes exemples afin de d'abord retrouver le verbe de la collocation, puis le nom, enfin la collocation entière (Montagnon G., 2007). Dans les trois cas, l'apprenant choisit la lexie dans un menu déroulant. En outre, une mention « Aide » (à côté de la case vide) donne la définition simplifiée de la lexie à insérer.
- (Figure 4) L'apprenant doit sélectionner les deux composants de la collocation, même (et surtout) si ceux-ci sont séparés d'une ou plusieurs lexies. L'exercice cible des passages longs, extraits de Scientext selon les critères sélectionnés par l'apprenant et la machine : la discipline et le champ sémantique (choix de l'apprenant), les extraits à éléments détachés (choix de l'outil (et du concepteur)).
- (Figure 5) L'apprenant doit situer le degré d'intensité des collocations de démarcation sur une échelle allant d'une intensité « partielle » (*interroger des résultats*) à une intensité dite « polémique » (*rejeter une hypothèse*). . Nous voulions aborder ces différences sémantiques importantes qui sont sujettes à confusion chez les apprenants peu habitués à se positionner dans leurs écrits scientifiques et donc peu enclins à l'utilisation du lexique approprié à cette fonction.

Pour chaque exercice, un retour (feedback) est fait selon la réponse donnée. L'apprenant peut toujours retenter une fois l'exercice s'il a fait des erreurs.



Figure 4 : Exercice de sélection des collocations de la séquence FULS

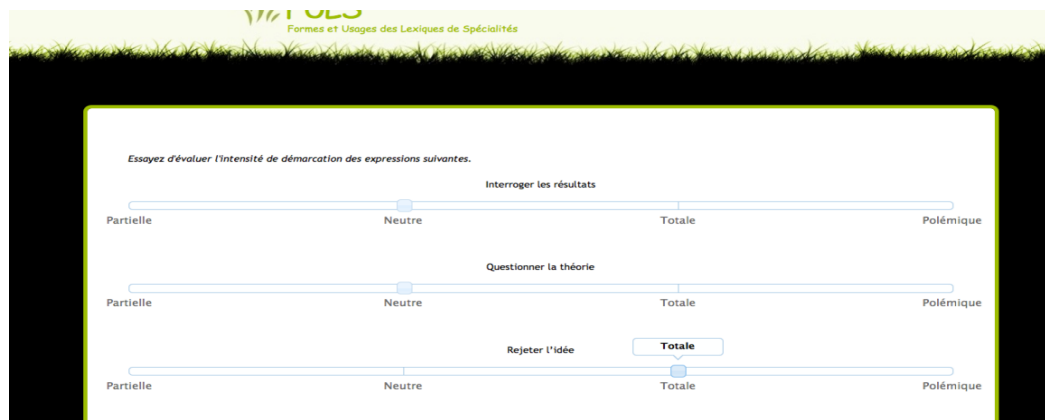


Figure 5 : Exercice d'évaluation de l'intensité des collocations

La suite de la séquence utilise des extraits plus longs (deux phrases) afin d'aider l'apprenant à cerner les contextes d'utilisation des collocations spécifiques. La séquence n'étant pas terminée en ligne, la suite est donc prospective. Nous voudrions en fin de séquence, placer

l'apprenant devant une partie triée du corpus pour une utilisation d'abord guidée puis individuelle. L'idée est de placer l'apprenant face au corpus mais de façon progressive. Le fait de proposer un corpus de petite taille et trié (Chambers A., 2005) permet de proposer, dans un premier temps, les exemples qui nous intéressent, sans « bruit ». Puis, nous souhaiterions laisser l'apprenant explorer Scientext en lui proposant des canevas de requête. La progression didactique envisagée conduit l'apprenant à faire des requêtes d'abord pour extraire un petit corpus (donc à l'aide de requêtes précises et simples) puis, pour extraire un corpus élargi en utilisant soit des requêtes plus complexes soit des requêtes déjà présentes dans Scientext (requêtes sémantiques préconstruites sur le thème du positionnement de l'auteur) qui permettent l'extraction de nombreux exemples. L'objectif final dépendra aussi fortement des visées de l'apprenant : soit apprendre à faire des requêtes, soit apprendre à trier pour analyser des grands corpus ; l'ordre de ces deux approches est lié à l'objectif de l'apprenant, mais les deux entrées sont à envisager pour des raisons différentes mais complémentaires.

...3.2.3 La séquence du côté des TICE

Avant de laisser l'apprenant rédiger seul son mémoire ou sa thèse, FULS finit sur une séquence d'aide à l'écrit scientifique grâce à un outil TICE « Scribplus¹² » (Figure 1 : « Réemploi en contexte de toutes les collocations abordées ») (Mangenot F. et Phoungsub M., 2010). Ce logiciel est une aide rédactionnelle pour les apprenants de FLE ; il permet – au départ – la rédaction d'écrits courts tels que des cartes postales ou des descriptions touristiques. Le logiciel Scribplus propose une série de questions qui accompagnent « progressivement l'apprenant dans la rédaction d'un texte, en l'obligeant à écrire des phrases à partir desquelles sera généré un texte définitif » (Phoungsub M., 2009): 3). Le principe se fonde sur une interactivité entre l'apprenant et le logiciel en associant les réponses de l'utilisateur aux questions/consignes du logiciel, et les énoncés proposés par l'outil. Ainsi, ScribPlus permet une certaine prise de conscience par l'apprenant des étapes de la production textuelle (Phoungsub M., 2011).

Précisons que Scribplus est la troisième génération de logiciel d'une famille née en 1992 sous le nom de « Scrivere con Word Prof » (Phoungsub M., 2009) en Italie, suivi de « Gamme d'écriture » (Mangenot F., 1998). L'objectif de FULS se situe dans l'utilisation du principe de ces logiciels d'aide à l'écriture qui ne nécessitent pas de développement informatique. Le principe qui nous intéresse, développé dans le travail de Tran (Tran T. T. H., 2010), est celui d'une aide à la production écrite scientifique et l'utilisation de la phraséologie appropriée. En outre, le double principe d'une aide des apprenants par le logiciel et l'enseignant, qui peut donner ses consignes et évaluer les productions des apprenants, est un compromis qui nous paraît pertinent pour un enseignement optimisé car complémentaire : les apports d'un outil (liste de collocations, automatisation de certaines chaînes de caractères...) et le savoir de l'enseignant. A l'aide d'une bibliothèque d'exemples de textes, d'un dictionnaire spécialisé et de consignes soit prédéfinies soit à concevoir, l'enseignant peut conduire l'apprenant vers les objectifs qu'il veut atteindre. L'outil serait alors programmé de façon générique pour l'aide à la rédaction d'une introduction et d'une conclusion d'un mémoire de Master (dans un premier temps) avec une insistance sur l'utilisation des collocations et de la phraséologie scientifique appropriée (Tran T.T.H., 2010 : 23 puis 34). Dans son travail, Tran (2010 : 47) a ainsi développé une séquence complète qu'il faudrait désormais implémenter dans Scribplus pour la tester.

Pour résumer, cet outil nous sert 1/ à placer l'apprenant face à son écrit et à l'aider à amorcer l'écriture en français, 2/ à l'aider à insérer dans ses écrits personnels les collocations

¹² Conception : F.Mangenot, Lidilem, Université Stendhal-Grenoble3, <http://scribplus.ovh.org/>
Réalisation : A.Khara, EIAL, Université Stendhal-Grenoble3

qui viennent d'être travaillées. L'apprenant ne va cependant pas pouvoir rédiger tout son mémoire de Master à l'aide de ScribPlus car l'outil n'est pas fait pour des écrits longs et cela semble – dans l'état actuel du logiciel – peu envisageable. Le choix d'une aide pour la rédaction de l'introduction et la conclusion s'est forgé au fil des analyses linguistiques faites sur le corpus. Notre choix s'est concentré sur ces sections car nous avons constaté que les collocations transdisciplinaires ont des fonctions rhétoriques parfois très précises au sein des écrits scientifiques. Ainsi, certaines ne sont utilisées que dans l'introduction ou la conclusion (Cavalla C. et Grossmann F., 2005, Cavalla C. et Tutin A., à paraître) ce qui – nous semble-t-il – peut faciliter l'aide à l'écriture pour des étudiants non natifs du français.

...3.3 Conclusion à propos de FULS

Pour conclure, il nous semble intéressant de résumer notre utilisation du corpus Scientext dans le projet FULS. Les didacticiens ne peuvent pas toujours utiliser directement les outils développés par les linguistes dans le cadre de la linguistique de corpus ; Au vu des contraintes d'utilisation dues soit à une terminologie linguistique non adaptée aux didacticiens-enseignants, soit à des modes d'interrogation du corpus qui correspondent plus aux attentes des linguistes qu'à celles des enseignants et des apprenants, FULS propose de guider les enseignants et les apprenants vers l'exploitation d'un corpus. Cette exploitation prévoit d'épargner à l'enseignant de passer du temps sur la création d'exercices à partir d'un corpus mais de le guider avec l'apprenant vers la découverte du corpus afin d'être finalement autonome dans l'exploitation de ce corpus.

Le corpus Scientext joue, dans ce cas, plusieurs rôles complémentaires : il permet la découverte d'éléments linguistiques spécifiques (collocations transdisciplinaires) ; il permet l'exploitation de textes authentiques très prisés dans l'enseignement des langues étrangères ; il permet le tri et le choix d'exemples spécifiques à différentes disciplines scientifiques. A l'aide de FULS, Scientext joue d'autres rôles directement pédagogiques : ses nombreux exemples triés servent à la mise en place d'exercices variés ; les phrases des exercices varieront, prochainement, à chaque nouveau démarrage de l'exercice ; l'extraction des collocations scientifiques du corpus a permis le développement d'une banque de données de ces collocations (surtout pour les introductions et les conclusions).

Ainsi, FULS permet d'exploiter un corpus selon deux entrées : d'abord l'entrée indirecte à travers les exercices élaborés à partir des textes du corpus, ce qui place l'apprenant face à des exemples authentiques mais choisis spécifiquement pour lui ; puis une entrée directe avec le développement futur d'une sorte de pilotage pédagogique conduisant l'utilisateur à interroger directement le corpus afin d'en extraire les données nécessaires. Il s'agit donc d'une aide à l'enseignant à la fois pour aborder un corpus et pour guider à son tour l'apprenant vers l'utilisation des outils d'interrogation de ce même corpus.

4 Bilan

Nous avons vu que les corpus sont peu utilisés en FLE malgré leur utilité, au moins au plan linguistique. Boulton (2008) propose que les manuels d'enseignement des langues étrangères intègrent des concordances comme exemples pour des exercices et permettent aux apprenants de travailler sur des éléments spécifiques. En effet, les manuels pourraient servir de tremplins vers l'utilisation des divers corpus existants en proposant des exercices à partir de concordanciers (par exemple). Les auteurs des manuels pourraient diriger les enseignants et les apprenants vers des outils gratuits en ligne qui proposent des concordanciers associés à des

corpus de langue générale (Lextutor¹³, Lexiquum¹⁴, Webcorp¹⁵) ou des corpus spécialisés comme Scientext. En attendant cela, nous proposons une séquence pour partie « clé en main », et pour une autre partie à construire en fonction des besoins linguistiques de l'apprenant. Nous adhérons aux préconisations de Granger (Granger S., 2001) et Loiseau (Loiseau M., 2009) qui soulignent l'intérêt du TAL qui permet à la fois de travailler sur l'objet linguistique et de créer une certaine genericité en vue de favoriser l'instrumentation. FULS n'est pas encore un tel outil, mais il propose des modules qui vont dans ce sens et des pistes pédagogiques réalisables dans cette perspective. Il constitue ainsi une sorte d'étude de faisabilité de nos activités didactiques autour du corpus Scientext.

Références

- ASTON G., "The Learner as Corpus Designer", KETTEMANN B. et MARKO G. (dir), *Teaching and learning by doing corpus analysis*, New York, The Edwin Mellen Press, 2002, p. 9-26.
- ASTON G., "Small and large corpora in language learning", LEWANDOWSKA-TOMASZCZYK B. et MELIA P. J. (dir), *PALC 97: Practical applications in language corpora*, Łódź, Pologne, Łódź University Press, 1997, p. 51-62.
- BINON J. et VERLINDE S., "Les collocations : clef de voûte de l'enseignement et de l'apprentissage du vocabulaire d'une langue étrangère ou seconde", *La lettre de l'AIRDF*, n°33, 2003, p. 31-36.
- BLANCHE-BENVENISTE C., "De l'utilité du corpus linguistique", *Revue Française de Linguistique Appliquée*, n°Numéro spécial: 'Corpus: de leur constitution à leur exploitation', Vol.1-2, 1996, p. 25-42.
- BLUMENTHAL P. et HAUSMANN F. J. E., "Collocations, corpus, dictionnaires", *Langue Française*, n°150, 2006, p.
- BOULTON A., "Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues", *Texte et Corpus*, n°3, 2008, p. 37-46.
- BOURIGAULT D., FABRE C., FREROT C., JACQUES M.-P. et OZDOWSKA S., "Syntex, analyseur syntaxique de corpus", (dir), *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles*, Dourdan, France, <http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=bourigault&subURL=bourigault-pub.html>, 2005, p.
- BRODINE R., "Integrating corpus work into an academic reading course", ASTON G. (dir), *Learning with corpora*, Houston, USA, Athelstan, 2001, p. 138-176.
- CAVALLA C., "Les collocations dans les écrits universitaires : un français spécifique pour les apprenants étrangers", BERTRAND O. et SCHAFFNER I. (dir), *Apprendre une langue de spécialité : enjeux culturels et linguistiques*, Paris, Editions Ecole Polytechnique, 2008, p. 93-104.
- CAVALLA C., "Méthodologie d'apprentissage de l'écrit universitaire", *Recherches et Applications*, n°47, 2010, p. 153-161.
- CAVALLA C. et GROSSMANN F., "Caractéristiques sémantiques de quelques 'Noms scientifiques' dans l'article de recherche en français", KIN T. (dir), *Akademisk prosa*, Bergen, Skrifter fra KIAP Romansk institutt, Universitetet i Bergen, 2005, p. 47-59.

¹³En ligne, gratuit : <http://www.lextutor.ca/concordancers/>, Compleat Lexical Tutor, Concordancier Corpus Français, UQAM (Université du Québec à Montréal), CSLP (Center for the Study of Learning and Performance).

¹⁴En ligne, gratuit : <http://retour.iro.umontreal.ca/cgi-bin/lexiquum>, Base de données textuelle, Université de Montréal, Rali (Recherche Appliquée en Linguistique Informatique), G.Lapalme.

¹⁵En ligne, gratuit : <http://www.webcorp.org.uk/live/>, Research and Development Unit for English Studies, Birmingham City University.

- CAVALLA C. et TUTIN A., "Etude des collocations évaluatives dans les écrits scientifiques", GAUTIER L. et MEJRI S. (dir), *Les collocations dans les discours spécialisés*, Dijon, Editions Universitaires de Dijon, à paraître, p.
- CHAMBERS A., "Integrating Corpus consultation in Language studies", *Language Learning & Technology*, n°9 - n°2, 2005, p. 111-125.
- CHAMBERS A., "L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé", *Revue française de linguistique appliquée*, n°XV, 2010/2, p. 9-20.
- CHAMBERS A., FARR F. et O'RIORDAN S., "Language teachers with corpora in mind: from starting steps to walking tall", *The Language Learning Journal*, n°39:1, 2011, p. 85-104.
- CHAMBERS A. et O'SULLIVAN I., "Corpus consultation and advanced learners' writing skills in French", *ReCALL*, n°16-1, 2004, p. 158-172.
- FLIGELSTONE S., "Some reflections on the question of teaching, from a corpus linguistics perspective", *ICAME journal*, n°17, 1993, p. 97-109.
- FLOWERDEW J., "Concordancing in language learning", *Perspectives*, n°5, 1993, p. 87-101.
- GRANGER S., "Didactique des langues étrangères, linguistique de corpus et traitement automatique des langues", *Questions d'épistémologie en didactique du français (langue maternelle, langue seconde, langue étrangère)*, 2001, p. 105-109.
- HABERT B., "Outiller la linguistique : de l'emprunt de techniques aux rencontres de savoirs", *Revue française de linguistique appliquée*, n°1 - Vol. IX, 2004, p. 5-24.
- JOHNS T., "Micro-concord: a language learner's research tool", *System*, n°14(2), 1986, p. 151-162.
- MANGENOT F., "Outils textuels pour l'apprentissage de l'écriture en L1 et en L2", SOUCHON M. (dir), *Pratiques discursives et acquisition des langues étrangères*, Besançon, Université de Franche-Comté, 1998, p. 515-525.
- MANGENOT F. et PHOUNGSUB M., "Un dispositif d'aide à la rédaction par incitations et socialisation", *Revue française de linguistique appliquée*, n°15 (2), 2010, p. 101-119.
- MANGIANTE J.-M. et PARPETTE C., *Le français sur objectif universitaire*, Grenoble, Presses universitaires de Grenoble, 2011.
- MANIEZ F., DURY P., ARLIN N. et ROUGEMONT C., *Corpus et dictionnaires de langues de spécialités*, Grenoble, Presses Universitaires de Grenoble, 2008.
- MCENERY T. et WILSON A., "Teaching and language corpora (TALC)", *ReCALL*, n°9(01), 1997, p. 5-14.
- OSBORNE J., "Phraseology effects as a trigger for errors in L2 English; the case of more advanced learners", GRANGER S. et MEUNIER F. (dir), *Phraseology in Foreign Language Learning and Teaching*, Amsterdam, John Benjamins, 2008, p. 67-83.
- PERY-WOODLEY M.-P., "Quels corpus pour quels traitements automatiques ?" *Traitement Automatique des Langues*, n°36, 1-2, 1995, p. 213-232.
- RABARDEL P., *Les hommes et les technologies, approche cognitive des instruments contemporains*, Paris, Armand Colin, 1995.
- REZEAU J., "L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne", *Alsic*, n°10/2, 2007, p. 27-43. Disponible en ligne: alsic.org.
- SCHAEFFER-LACROIX E., "Exploitation de corpus lors de la production d'écrits électroniques en L2", *Revue électronique Texte et corpus*, n°3 / août. Disponible en ligne: http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_schaeffer_lacroix.pdf, 2008, p. 47-56.

TRAN T.T.H., *Aides logicielles à la rédaction universitaire. L'exemple des collocations*, Grenoble, Université de Grenoble, Mémoire Master Recherche FLE, URL: dumas.ccsd.cnrs.fr/dumas-00563974/en/, 2010.

TUTIN A., "Collocations du lexique transdisciplinaire des écrits scientifiques : annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique", *Cahiers de l'Institut de linguistique de Louvain*, n°Vol. 31, n° 2-4, 2007, p. 247-262.

TUTIN A., "Modélisation linguistique et annotation des collocations : une application au lexique transdisciplinaire des écrits scientifiques", KOEVA S., MAUREL D. et SILBERZTEIN M. (dir), *Formaliser les langues avec l'ordinateur : de Intex à NooJ*, Besançon, Presses Universitaires de Franche Comté, 2007, p. 189-215.

TUTIN A. et GROSSMANN F., "Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif", *Revue Française de Linguistique Appliquée*, n°VII, 2002, p. 7-25.

VERLINDE S., BINON J. et SELVA T., "Corpus, collocations et dictionnaires d'apprentissage", *Langue Française*, n°150, 2006, p. 84-98.

VERLINDE S., SELVA T. et BINON J., "La base lexicale du français: de la lexicographie d'apprentissage à l'environnement d'apprentissage", HEINZ M. (dir), *Le dictionnaire maître de langue. Lexicographie et didactique*, Berlin, Franck et Timme, 2009, p. 289-306.